Uses and Misuses of the Automatic Identification System

Clément Iphar, Cyril Ray, Aldo Napoli

IEEE OCEANS Conference - Marseille, France - June 17-20 2019 - doi: 10.1109/OCEANSE.2019.8867559

Uses and Misuses of the Automatic Identification System

Clément Iphar
CRC - MINES ParisTech
PSL Research University
Sophia Antipolis, France
clement.iphar@mines-paristech.fr

Cyril Ray
IRENav - French Naval
Academy Research Institute
Brest, France
cyril.ray@ecole-navale.fr

Aldo Napoli
CRC - MINES ParisTech
PSL Research University
Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

***Abstract*—The Automatic Identification System (AIS) is widely used by mariners at sea as a compulsory collision prevention device on-board. Since the inception of collaborative platforms that gather AIS messages received by a network of stations, large worldwide AIS datasets can be assembled and enable a series of analyses of the maritime traffic, including an increase of the maritime domain awareness, the assessment and prediction of trajectories, the detection of anomalies, the measurement of fishing pressure, vessel-born noise and air pollution and the global modelling of the maritime traffic. However a series of misuses of the AIS does exist, and data may lack reliability as some users broadcast errors or intentionally falsify the information sent by the device. In this respect, a methodology to assess the integrity and veracity of AIS messages has been developed, leading to the determination of maritime risks in support to maritime monitoring activities. In this paper, an exemplification is proposed, based on AIS possible integrity issues on kinematic data.**

***Index Terms*—Automatic Identification System ; data falsification ; risk analysis**

I. Introduction

The ocean, crossroads of international issues, is facing a growing pressure of human activities. In some domains such as goods transportation or energy transportation, up to 90% of the world traffic is done by sea. Fishing, sailing and cruising are amongst the domains of maritime activities, generating an important and ever-increasing traffic. As the number of vessels using these routes, entering or leaving ports and crossing heavily occupied areas increases, navigation difficulties arise. In addition, the high number of sailing vessels, each one exhibiting its own movement with its own objective, may lead to conflicting situations, which may evolve into hazards for maritime navigation.

In order to enhance the security and safety of navigation, several systems have been put in place by coastal states or international organisations. Those systems are either active (such as Vessel Management System) or passive (such as radar). Their purpose is to localise vessels and give coastal states and vessels at sea a genuine overview or the surroundings. One of those systems is the Automatic Identification System (AIS), in which vessels broadcast localisation messages to other vessels and coastal stations in their neighbourhoods.

The AIS has numerous weaknesses and shortcomings as the transmission is neither secured nor verified. Its intrinsic weaknesses cover the issues of missing data, message collision and errors within the data. Some AIS transponders fail to reach all the requirements set by the International Telecommunications Union, and some ships display large blank areas. This missing data, as shown in [1], weakens the exploitation of AIS data by decreasing the reliability, but does not prevent it (as a meaningful statistical study is needed in order to judge the quality of data). The AIS has some critical shortfalls in additions to problems such as limited bandwidth and range: limited retransmit capabilities for a few messages and no retransmit capabilities for the majority [2]. There is no way to detect dropped packets (typically observed by embedding sequence number in packets), there is no mechanism to verify the identity of the sender, so a large portion of the message may be problematic.

Message collision can have a variety of causes: the protocol has shortcomings, using the principle of local self-organisation, resulting in possible collision as vessels do move ; message overlapping which happens when two messages are sent during two adjacent time slots and overlapping due to the path difference of the signals (it often happens in satellite reception, as the AIS has only a 12 bits delay and was not designed to handle high height reception). [3] estimated that for a channel load of 20%, 5% of messages were lost, for a channel load of 60%, 15% of the messages were lost and 25% of the messages were lost with a channel load of 90%.

Some data that AIS messages contain are entered manually at the system initialisation (first use). A study of errors in those data is done by [4]. These errors are not intentional, they can be done by underestimating the importance of a proper fulfillment of the fields or by ignorance of the way the system works. For the unique identification number of the ship (also called MMSI number), 2% appear to be erroneous, and 6% of the vessels have no specified vessel type. In the dynamic data, errors are also widespread, as stated by [4]: as for the position, 1% of vessels have a latitude greater than 90 or a longitude greater than 180; at least 30% have an erroneous navigation status, e.g. say they are sailing while having a discordant speed, and the destination is unclear (vague name, country, abbreviation) in about half the cases.

This paper is structured as follow: Section II first lists a non-exhaustive but wide range of uses of the AIS, then Section III surveys the various known misuses of the system, including the exploitation of the system intrinsic weaknesses and the input of falsified or spoofed information. After the presentation of the uses and misuses, the methodology we propose for message assessment is presented in Section IV, before the determination of maritime risks of a selection of maritime scenarios based on integrity issues of AIS kinematic information in Section V.

II. The uses of AIS

AIS is widely used for drafting a picture of the maritime situation at a given time. Due to the large number of vessels carrying it, it offers a wide and somewhat truthful view of the state of the marine traffic. Some of the main maritime situation applicative domains for the use of AIS messages which are maritime domain awareness, anomaly detection, trajectory analysis, vessel prediction and data fusion are presented in this section. On top of that, AIS data is widely used for various applicative models of the maritime network and the maritime navigation. AIS, given its high frequency and carriage obligation for a large number of vessels, provides a reliable picture of the maritime navigation, at least for the vessels obliged to carry the system. In addition, AIS provides a wide range of data, which can be used in a wide range of applications.

In general, it must be noticed that most of the studies take the AIS data “as if” and do not question their genuineness, in particular the presence of system issues and errors (as presented in Section I) and numerous falsification possibilities and spoofing cases (as presented later in Section III). Some studies, such as [5] or [6] even create missing data from neighbouring data, in accordance with their need for particular data at a given time.

A. Maritime domain awareness

Maritime Domain Awareness (MDA) is the detection, classification, identification and monitoring of vessel data in order to understand the activities at sea [7].

The needs are to check that the passage in the territorial waters is harmless (in accordance to maritime laws), so in order to fulfil those needs control bodies are used, to which analysis tools must be provided for a spatio-temporal monitoring of sea activities. The concerns of the administrations in charge of the maritime domain awareness are numerous [8]: prevention of accidents, particularly on ships which carry hazardous material ; detection of oil spills and generation of alerts for agencies in charge of oil spill operations ; support anti-piracy operations; monitor the maritime borders ; monitor the fisheries, and detect illegal, unregulated and unreported fishing ; the detection of illegal trafficking and smuggling ; the support of authorities in S&R (Search and Rescue) operations. Thus the stakes are maritime security: port security, container tracking, effective S&R, and maritime safety: struggle against trafficking, piracy, smuggling, illegal immigration, illegal pollution or antiterrorism [9] [10].

The Maritime Situational Picture is basically the maps of the locations of the ships in a given area of interest, at a given time. A proper Maritime Situational Awareness requires the ongoing maintenance of the Maritime Situational Picture [11]. In this scope, the goals of AIS for Maritime Domain Awareness are the coverage, the network, the interoperability and the data management [12]. To fulfil those goals, the system must be made usable for Maritime Domain Awareness by the validation of data, correlation, data fusion and the storage of AIS data in an usable way for future analysis. The policy issues around AIS are numerous, for the management of binary information (that must not affect current and future AIS equipment on-board ship by causing mariners to stop their use of the system), the data sharing policies, the use of AIS for other purpose than its original ones (that could cause people to be reluctant in its use), the frequency allocation and the use of other frequencies in some parts of the world (which prevents the satellite system from receiving all data), and the enforcement of carriage [12].

B. Collision risk assessment

The collision risk is the main application of this section, as some waterways are busy, such as Rotterdam (30,000 sea-going ships and 135,000 inland vessels per year) or Shenzhen, near Hong Kong, with 500,000 ships per year. In the down-stream Yangtze, up to 5,000 ships are transiting daily. The collision occurrences would climb proportionally with the number of ships without traffic management and traffic services [13]. In European waters, some disasters such as *Erika* (December 1999), *Ievoli Sun* (October 2000), *Prestige* (November 2002), *Tricolor* (December 2002) occurred, underlining the necessity of collision risk reduction.

Collision risk assessments use the notion of ship domain, which is the immediate surrounding waters of a vessel, that a navigator would want to keep clear of any other ship. Regulations, such as COLREGS, are a major point of collision risk prevention [14], [15]. Collision reasons are various and research has concentrated on collision probability computations [16].

C. Anomaly detection

Detecting and classifying abnormal behaviours is a key task of maritime situational awareness, for several reasons such as the extraction of relevant contextual information and the proper monitoring of both self-reporting systems (such as AIS) and non-cooperative systems (such as satellite imagery or coastal radar). The operator must get information with a quality which is good enough to make a decision but also to understand the underlying meaning of the data handled, through evaluation criteria. Those criteria are numerous, but the main ones are uncertainty, imprecision and trueness [17], the uncertainty being the degree of confidence assigned to a specific value (when one is known to be true), and can be caused by lack of knowledge (epistemic uncertainty) or random variability of the process (aleatory uncertainty); imprecision being the inability of the source to provide a single value or to discriminate between several values; and the trueness being the criterion linking a piece of information to the truth (or reference), also referred as the closeness of agreement between the expectation and the measurement.

The different kind of anomalies that can be met are the kinematic (anomaly shown in the motion of the vessel) and static (anomaly shown in the properties of the vessel) [18]. They can be further divided in subgroups, such as manoeuvring issues (involving the velocity vector of a vessel), location issues, interaction issues (illicit or unusual interaction between vessels, or between a vessel and an infrastructure) for kinematic anomalies. The detection of anomalies can then embrace the cases of vessel stopping, vessel loitering, the entrance in an exclusion zone, the crossing of an exclusion zone, or a rendezvous detection, amongst others.

D. Trajectory analysis

The analysis of trajectories are a main feature of AIS data. [19] define a trajectory as "*a record of the evolution of the position (perceived as a point) of an object that is moving in space during a given time interval in order to achieve a given goal*". The AIS, in this scope, provides data at a high reporting rate enabling the analysis of trajectories, provided that a sufficient amount of messages is received by the receivers.

For reliability assessment of data suspected from being incomplete, on the one hand, for each pair of message, the time elapsed between the messages can be compared to the expected time between two messages (which depends on the speed of the vessel). On the other hand, the position of the second point can be calculated with respect to the position, the speed and the heading of the first point in order to assess whether or not the segment is reliable. In order to do so, [1] proposed a method in which a statistical distribution of positions was put in place and a risk value assigned to the points of the trajectories, leading to the determination of the lack of reliability of such a segment using Shannon's amount of information for the determination of the occurrence of events.

Spatial aggregation of trajectories based on AIS data is discussed in [20], as discrete data is not always usable for analysis purposes, as well as incomplete data. Two approaches are then usable when it comes to spatial aggregation of trajectories: discrete and continuous. Continuous approach preserves spatial pattern as they avoid distortion due to discretisation of space, and discrete approach gives accurate numeric measures of movements.

E. Vessel prediction

The best way to track a vessel and predict its future position is to rely on the way it moves. Different methods for vessel prediction can be used, such as point-based, acceleration-based, heading-based, vector-based or cog-based (based on the course over ground). *Ad hoc* systems [21] compute predicted positions with at least two actual updated positions, and then when a new piece of information is provided, it can be compared to the predicted position. If the difference is larger than a given threshold, new updated positions are used for a new prediction.

The point-based prediction can be used when the ship does not move a lot, *i.e.* when anchored or moored. The vector-based approach can be used when the vessel has a longtime linear and constant speed. At the beginning of such a phase, cog-based should be more efficient, then heading-based before the eventual vector-based. When the vessel accelerates or decelerates, acceleration-based predictions shall be used.

F. Data fusion

The purposes of data fusion are to ingest data from different sources, to track vessels from certain areas in both real-time and offline cases, to provide a maritime situational picture and to associate cooperative reporting systems with non-cooperatively detected vessel [22]. None of the sensors such as AIS or radar can provide sufficient data on a regular basis for the establishment of a accurate and reliable picture of the maritime traffic at each moment. Radar is less accurate than AIS, and is subject to weather conditions, while AIS is fully dependant on its cooperative nature [23]. The fusion of data can improve the quality of information.

Numerous methods to deal with multiple data of the same type exist, however few exist to deal with multiple data of different types. The multi-type multi-source data fusion is complicated and computationally expensive. Single-type multi-source data fusion techniques are Neural Network learning, Bayesian learning, Kalman filtering and Dempster-Shafer evidential reasoning.

G. Fisheries

Purpose of using automatic identification of maritime activities is the improvement of maritime situational awareness and is to fight against Illegal, Unreported and Unregulated (IUU) fishing, which destroys marine habitats, depletes fish stocks, put honest fishermen at an unfair disadvantage, distorts competition and weaken coastal communities [24]. As the future economic viability of fisheries and the biodiversity need sustainable stocks, a data processing is necessary.

In [25] the study of the match between the EU fishing fleet register and an AIS dataset was performed, with a 70% rate of matching. The main causes of mismatch were the absence of call sign identifier in the EU fishing fleet register (3.8%), the misspelling of the identifier or the fact that the vessel was not present in the AIS dataset (because if the lack of coverage, the fact that the vessel is not operative, the fact that AIS is not used).

Besides this study, [26] used 5 years of satellite data (January 2011 to October 2015), for the identification of fishing patterns: three dominant fishing gear types are discriminated: trawl, longline and purse seine. It must be noticed that [26] is aware of the fact that come people actually falsify positional or identification data.

H. Traffic modelling

The modelling of traffic can also benefit from AIS data, as it provides navigation patterns. Before access to a large amount of vessel data, models were made on small samples [27], but now with AIS, large quantities of vessel data can be processed. Straits and port entrances are particularly hazardous locations due to the bottleneck effect of traffic. Measures can be taken, such as TSSs, but only 15 straits have implemented an IMO-approved TSS so far [27].

AIS is a useful tool to gather information for sea traffic control [28], and the characteristics of the traffic are important elements to be known as they provide information on the main routes, the relative use and the pressure on some areas. Data Mining techniques can be used to learn characteristics of the vessel traffic [29], and clustering analysis was performed by [30] in the Shanghai port. The statistics can include the number of vessels following the same route, the width of the route, the legs of the route, or information relative to the speed and structure of the ships on this route [31]. For a quantitative qualification of security, the realisation of diagrams of ship movement intensity can be performed [32], and algorithms for vessel collision probability has been set up by [31], using navigational situation data such as area, hydrological and meteorological conditions data, the own vessel motion parameters and maritime navigation rules and regulations.

I. Vessel noise

Noise exposure from human activities can take several forms, such as seismic surveys, pile-diving operations, military sonar activities, and vessel navigation noise [33]. The AIS data can then help in assessing anthropologic components of the underwater noise, and on the other side noise measurements can help in the quantification of human activities in the maritime domain. However, as all vessels are not required to carry an AIS receiver, the use of AIS is not perfect for noise determination [34], [35]. The anthropogenic marine noise have deleterious effects on marine organisms such as mammals, fish and cephalopods, as such noise can mask naturel means of communication [36], or chronic stress [33].

J. Vessel emission

Maritime transportation makes a large contribution to air pollution and greenhouse gases emission. An estimation of such emissions can be performed on both quantification and location. Several different of data must be collected for an emission evaluation of vessels, such as technical information on vessels, specific consumption and emission factors, ship activities and maritime routes. For ship activities, AIS is the system which, when used for a certain period, offers more accurate estimation of vessel activities, improved reliability of fuel consumption and emission estimations [37]. However, the use of AIS can raise some problems, such as possible low penetration of AIS in some areas and beyond network coverage and suitability of AIS for the air estimation around ports. Indeed, the discrepancies in coverage according to the areas create a misleading picture of maritime traffic. In air management policies, marine emissions are often neglected, but they should not [38], as adverse effects of emissions have been demonstrated on human health [39], with a significant impact for PMs (Particulate Matter), and the social cost of ship emission is evaluated by [40] at an estimated 44 million dollars for each billion dollars of port revenues.

The various gases taken into consideration are both greenhouse gases and pollutants and are carbon monoxide (CO), carbon dioxide (CO_2), sulphur monoxide (SO), sulphur dioxide (SO_2), Nitrogen oxides (NO_x), little particulate matter ($PM2.5$), big particulate matter ($PM10$), ozone (O_3), methane (CH_4), nitrous oxide (N_2O), polycyclic aromatic hydrocarbon (PAH), volatile organic compounds (VOC), ammonia (NH_3) and hydrocarbons (HC).

K. Animal collision

AIS data is used in the case of collision between vessels and animals, particularly marine mammals. Indeed, some routes cross areas frequented by marine mammals [2], and the fact that the traffic increases is a problem in the scope of the whale-ship collisions, which is a mortality cause of significant importance for endangered whales [41].

Off the coast of the Massachusetts state of the USA lies the Stellwagen Bank National Marine Sanctuary. At is covers most of the traffic entering or exiting the Massachusetts Bay and the Boston harbour, it undergoes heavy traffic. In [41], AIS data from 2006 coming from this area was studied, consisting in over 2 million points from over 500 vessels. Using a model for death probability of whales, [41] implemented speed limits in order to assess the risk mitigation of a speed reduction, by setting all over-threshold speed data to the selected threshold. Significant differences were assessed, as when the threshold was put at the mean transit speed (13.5kn), the lethality of the collision was 67%, while it was 96% at 16kn and 43% at 10kn.

III. The misuses of AIS

A. Data falsification within the AIS

Intentional falsification of the AIS signal is done by the crews on board the ships in order to modify or stop the message they send, in the very particular purpose of misleading the outside world. AIS data falsification covers several aspects, presented hereafter.

Identity theft also exists in the maritime domain [42]. It corresponds to the fact to navigate with a MMSI (Maritime Mobile Service Identity) number which is not the real one, allocated and internationally recognised, but with the one of another vessel that actually exists somewhere else. Hundreds of ships are disguised this way. As the MMSI number changes, there is no way to assess *a priori* whether the vessel one is looking at is the right one.

The problems met with AIS are numerous, and most of them are the consequence of the non-secured nature of the transmission. However, the fixes that would be needed to have a reliable AIS system are at the protocol level. Five main issues are developed by [42]: the identity fraud, the concealing of destination, the fact to voluntarily stop the broadcast, the GNSS manipulation and the spoofing of the system.

Identity fraud is the fact to use a false or stolen identifying mark of a vessel. An estimated 1% of the vessels assessed by [42] used a false IMO (International Maritime Organization) number. The impact is then that anyone interested in AIS data has no assurance that the name displayed corresponds to the selected vessel of interest.

Concealing the destination occurs more than half of the time, when vessels do not report their next port of call (41% rate [42]). The impact is the creation of information gap (as it is not possible to know when the vessel will arrive) and possibly an intentional mislead, a skew of the view of the state of the traffic.

The fact to "go dark" is the fact to turn off the AIS transmission, as over a quarter of the vessel turn off their system at least 10% of the time (once the active shut downs and lack of satellite coverage taken into consideration). As the large vessels (over 250m in length) are more likely to turn off their transmission, this is suggesting that they have greater will to conceal some of their activities such as fishing in an unauthorised area, or trade illegal goods [43] with other ships or on coasts. The impact is that it undermines the ability to track vessels, and it is challenging for financial and security stakeholders.

As there is no validation of GNSS location, the coordinates can be changed. From mid- 2013 to mid-2014, a 59% increase in GPS manipulation has been observed [42]. The impact is the difficulty it brings to know the actual location of a vessel.

B. AIS spoofing

The spoofing of messages is done by an external actor by the creation *ex nihilo* of false messages and their broadcast on the AIS frequencies [44]. Those spoofing activities are done in order to mislead both the outer world and the crews at sea, by the creation of ghost vessels, of false closest point of approach trigger, a false emergency message or even a false cape (in the case of a spoofed vessel).

In the scope of spoofing capabilities, several threats can be taken into consideration: ship spoofing, aid to navigation spoofing, collision spoofing, AIS-SART spoofing, weather forecasting, AIS hijacking and availability disruption threats [44].

Ship spoofing consists in the crafting of a valid non-existent ship, with the assignation to the ship of false static information (this can be based on the duplication of real ship information). A wide range of malicious scenarios can be imagined, such as spoofing a vessel into the jurisdiction water of an enemy, making a nuclear carrying vessel sailing in the waters of a nuclear-free nation, amongst others. An attacker would also be able to counterfeit information to blame someone else about an event, for instance a voluntary oil spill in the open sea.

Aid to navigation spoofing consists in the crafting of false data to lure a targeted ship into a manoeuvre that could be wrong and possibly hazardous. For instance it would be the fact to place buoys at the entrance of a harbour, or to place buoy to instruct a vessel to navigate in low waters.

Collision spoofing is the fact to create a ghost vessel which would cross the trajectory of the targeted vessel and trigger a CPA (Closest Point of Approach) alert, which could lead the vessel off-course, possibly running aground or into a rock.

AIS-SART spoofing consists in the generation of a false distress beacon for man overboard at given coordinates in order to lure and possibly force the target vessel into an attacker-controlled area, as by law a vessel is required to join an ongoing rescue operation upon the reception of such message.

Weather forecasting can be involved in the case of spoofing of binary messages, which convey messages such as weather and in this threat false weather forecast can be done and sent to the vessels for instance to influence optimal ship routes.

AIS hijacking consists in the alteration of information of existing AIS stations, with eavesdropping on the communication and replacement of some AIS data. The recipient receives a message which is not the one sent, as the attacker overrides the original transmission by broadcasting the fake signal with a higher power.

Availability disruption threats are three of a kind: slot starvation, frequency hopping and timing attacks. Slot starvation consists in impersonating the maritime authority to reserve all the spots, thus all stations within coverage have no slot available for reservation and emission. Frequency hopping in the fact to instruct the AIS transceivers to change their transmission frequency, as it is possible by protocol specification for given areas in the World. In timing attacks, the malicious user instruct transceivers to delay their transmission, by doing it repetitively, it prevents the system from functioning normally; and on the contrary, the attacker can command transceivers to send updates at a very high rate, thus overloading the channel.

C. Implications of AIS misuses

The implication of a false, falsified or spoofed AIS are, first, on the subject of the safety of maritime navigation, as it weakens the view one can have of the marine traffic, of its surroundings for a vessel, off its coasts for a country or of its fleet for a company. But there is also an implication for finance as it brings a distorted view of the flows of goods, a flawed understanding of supply and demand, and an impact on trading models [42].

For security and law enforcement, the main implications are the fact to trust no one, as AIS data cannot be fully trusted as it is manipulated. The use of AIS for maritime control requires the ability to assess AIS data; the existence of ghost ships, potentially elevating political tensions in case of malicious vessel appearance; the erasing of footprints, by the removal of tracking data of the activities of the vessel; and the fact to undermine watch lists as, by concealing their activities and their identities, vessel avoid the watch lists held by ports and authorities, dramatically decreasing their effectiveness.

IV. AIS MESSAGES ASSESSMENT

A. Methodology

Due to the vulnerabilities of the Automatic Identification System, and since an increase of the risk of maritime navigation is a consequence of those shortcomings, we propose a method to assess those risks. This method, based on an integrity assessment, is thoroughly presented in [45]. A detailed study of the AIS leaded to the determination of 935 unique integrity assessment items, taking into consideration the complex structure of the system, each of those items, properly uniquely and unambiguously denoted by an *ad hoc* nomenclature, representing one case of possible integrity failure within AIS data.

Each assessment item is an elementary case in which AIS data could not be in accordance with what is expected from the technical specifications point of view, or display an integrity issue of the system by transmitting two or more incoherent pieces of information. Four assessment steps have been set: a first step in which a data field from one message is taken individually and treated with respect to the system specifications, a second step in which several data fields of one single AIS message are compared, a third step in which a single data field or several data fields from several messages of the same type (for instance a series of scheduled position report messages, number 1) are compared and a fourth step in which data fields from several messages of various types are assessed. Groups of items can be established on the basis of their main purposes, that can be conformity issues, incoherent data field values, field value evolution inconsistency, trajectory issues, unusual values, an infrequently high rate of transmission, a transmission between vessels supposedly too remote to communicate, an unexpected change of the value of a fixed data field, issues linked to the positioning or an incoherent answer to a request. For each message, every single item is rigorously assessed following a logic-based frame and a value *True* is assigned to the item if it does display an integrity issue, else the *False* value is assigned.

The fact to use only AIS data would be sufficient in the case of an isolated system. However, the vessels evolve in a complex environment that must be taken into consideration in the frame of an integrity-based study of AIS, particularly when we want an assessment of associated risks as an outcome. Thus, every single dataset from which an attribute can be compared to one or several AIS data fields is of interest for this complementary assessment. The AIS covers a large spectrum of information, and various datasets can be used in this respect. Three main families have been discriminated: the environmental datasets (providing information on the environment in which the vessel evolves, such as meteorological data), the vessel-oriented datasets (including for instance fleet registers) and the navigation-oriented datasets (with the geometries of areas of interest such as traffic separation schemes or anchorage areas).

In order to enable a description of the situation of the vessel when the message was sent, a system of *flags* has been set, a

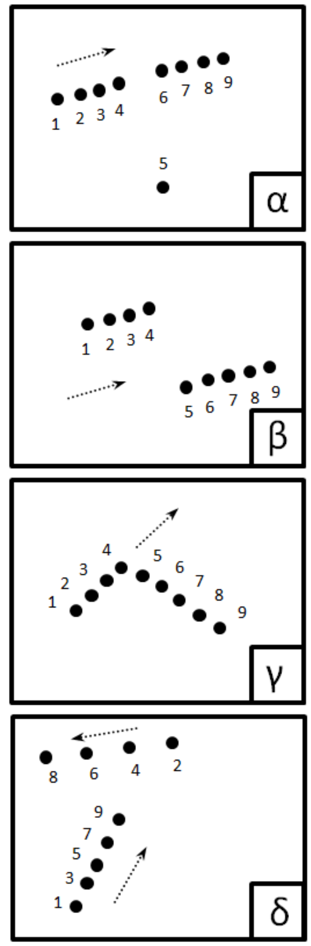

Fig. 1. Various considered scenarios and numbering of AIS contacts within those scenarios

flag being a simple element in natural language describing the instantaneous state of the vessel, as a consequence of the series of analyses this message underwent. A flag takes a Boolean value of *True* if the described situation is realised or *False* otherwise.

B. Use case: Kinematic inaccuracies

Kinematic inaccuracies are issues in AIS messages directly linked to the kinematic data fields within data fields, which are the latitude, the longitude, the speed over ground, the course over ground and the rate of turn. The knowledge of those data fields enable a reconstruction of the trajectory of the vessel, a prediction of the future locations of a vessel or the assessment of the genuineness of a message using the kinematic values harvested in previous (last received) messages.

Four cases of possible kinematic issues have been discriminated in our study and are presented in Figure 1. In this schema, each point stands for an AIS contact, and a total of 9 contacts are presented. Later, in Section IV-D, a set of 9 AIS messages will be synthetised following this pattern, in order to provide a testing ground to the algorithms.

The case α is one single outlier point out of a regular-looking trajectory, that can originate from an error of the system, a falsification test at the level of the station or a spoofing test from an external actor. The case β shows the brutal shift of a trajectory which seems to be regular before and after it. Such a case can be caused by a continuous software or hardware malfunction in which a bias has been activated, a proper falsification from the vessel or a spoofing case. The falsification case could be favoured if the area in question is in the near location of a place where a vessel could be willing to hide some information. The case γ draws the case of a vessel changing its course in a regular-looking trajectory, but in which other AIS data would disagree with the observed change. The same causes as case β can be proposed. The case δ displays a trajectory of a vessel going back and forth from two locations (more or less remote), actually forming two separate trajectories of two vessels sailing under the same identification number. This case would be typical of identity theft or identity sharing.

C. Flags of interest

Amongst the set of flags defined and implemented, a subset of three flags of interest in the case of kinematic issues can be isolated. This section presents the main characteristics of each of those flags.

1) Flag "f_nextpos": The flag standing for issues on two consecutive positions is called *f_nextpos*. Its computation is only based on two consecutive AIS messages. Using the timestamp, the position, the course over ground, the speed and the rate of turn of the vessel in the first message, an expected location can be computed for the second message once it has been received, using its timespan. If the actual position of the vessels differs from the computed one in a magnitude superior to a given threshold, the flag *f_nextpos* is raised.

2) Flag "f_trajectory": The flag standing for issues on a full vessel trajectory is called *"f_trajectory"*. Its computation is based on several consecutive messages, at least 5, sent by the same vessel. The vectors of timestamps, positions, courses over ground, speeds over ground and rates of turn are processed in order to find out whether or not the trajectory is consistent as a whole. If, out of the studied sub-trajectory, at least one point is found out to be an outlier (*i.e.* to divert over a given threshold from the mean trajectory), the flag *"f_trajectory"* is raised.

3) Flag "f_ubiquity": The flag standing for vessels displaying ubiquity issues is *"f_ubiquity"*. Its computation is only based on two consecutive messages. Using only the timestamps and positions of both messages, the temporal and physical distances between the two messages are computed. Then those distances are compared to a threshold speed which represents the maximum expected speed for this vessel. As a consequence, if the computed speed is above the threshold, the flag *"f_ubiquity"* is raised, and it means that both messages were not transmitted by the same vessel, and therefore that at least one of those two vessels broadcast its information under a falsified identity.

D. Synthetic data generation

This methodology is supported by a self-harvested AIS dataset of 6 months of data, our antenna being located in the Brest roadstead, France. This dataset is presented in [46]. In complement to the regular AIS data, and in order to comply with the specificities of the scenarios described in Section IV-B, several sets of synthetic data have been produced. Each synthetic dataset is composed of nine AIS contacts, mirroring the data points presented in Figure 1. The nine points are numbered from 1 to 9 and temporally sorted, so that their succession forms the wanted coherent trajectory. Then, those specifically synthetised datasets are included in the original AIS dataset, and are expected to trigger the flags presented in Section IV-C where those specific messages are treated by the system.

E. Risk association

Maritime risks are various and we classify them in four categories: natural, anthropic, environmental or maritime. Natural risks can be predicted such as storms or unpredictable such as tidal waves or collisions with cetaceans. Anthropic risks are linked to the human use of the oceans, in particular the fact that a fair number of vessels sail under a convenience flag. Environmental risks gather the risks linked to be far away from the coasts (such as diseases) and to carry hazardous material (such as oil spills). Last, the so-called maritime risks are directly linked to the presence of other vessels, such as collision risk or piracy risk.

In the frame of our study, five main risk families have been selected, as a consequence of their fitness to the issues linked to the AIS: the collision and boarding risk, the grounding risk, the illegal fishing risk, the piracy and terrorism risk and the risk of illegal transportation of goods or people. In order to link the selected risk families and the AIS issues highlighted by the data processing, the establishment of typologies [47] describing the environment of maritime traffic and an ontologic frame linking the stakes, the actors, the anomalies and the maritime environment in general.

In order to link the flags of the message analysis to the risks, it is necessary to perform a flag-based risk analysis, the flags being considered either alone or in groups. Thus, a list of flag activation scenarios has been set, each of those scenarios consisting, for each message, in the activation of one or several flags. To every of those scenarios is associated one or several of the risks taken in the list of the five selected families, and when all the flags associated with a scenario are selected, the scenario is considered as realised and the corresponding risks are activated. If several scenarios are activated, their risks are aggregated. In this study, a table of correspondence between the scenarios, the flags and the risks is needed, and has been realised and implemented with the support of maritime domain experts.

In order to decide on a risk level to be assigned, several elements have been taken into consideration: the type of the vessel, the dimensions of the risk and the type of risk. The vessel types we consider are: cargo vessels, cargo vessels carrying hazardous goods, passenger vessels and a last category gathering pleasure crafting, fishing and service vessels. The three dimensions of the risks taken into consideration are the human dimension, linked to the endangering of the human life at sea, the infrastructural dimension, linked to possible damages underwent by vessels, off-shore structures or ports and the environmental dimension, including oil spills. Risks levels are four: minor, moderate, major and extreme. Thus, for each of the selected risks family, a table giving the level of risk to be applied with respect to the vessel type and to each of the three dimensions of the risk have been set with the support of maritime domain experts. The study of flags provided us a list of risks to assess, and this assessment is performed using those tables.

The main drawback of this risk assessment is the deterministic nature of the study. Indeed, a deterministic approach has been preferred as it better fits our proof of concept and its moderate implementation complexity. Albeit imperfect, this approach enables risk and anomaly detection, and risk levels assignments, following a set of expert rules. However, this approach does not allow to take fuzziness into consideration, which is an important feature of risk analysis. An inference engine could orientate this study towards a statistical approach, taking into account the links between typologies that have already been individuated.

V. Flag computation and risk assessment

A piece of software written in python for functions and in SQL for queries, supported by a Postgres/Postgis database of AIS and contextual data, has been developed, with ca. 650 items and over 30 flags implemented and usable on a large scale dataset. In-depth implementation details on all aspects of the computation within the piece of software can be found in [45].

This section presents the results of the computation for each of the four cases presented in Figure 1. Table I displays the outcome of the case of a single outlier (case α), table II displays the outcome of the case of a trajectory shift (case β), table III displays the outcome of the case of a change of course (case γ) and table IV displays the outcome of the case of ubiquity (case δ).

In each table, the columns numbered from 1 to 9 represent the data points generated by data synthesis and isolated from the rest of the dataset, their numbering corresponding to the numbering of points in Figure 1. In the upper part of each table one can find three lines for the three flags, and the flag is raised for the given data point if the symbol $\surd$ is present in the box, else the flag is not raised. In the lower part of each table one can find five lines standing for the five selected risk types, the risk is considered for the corresponding point if the symbol $\surd$ is present in the box, else the risk is not activated. The lower part is generated in a deterministic way from the upper part following expert rules presented in section IV-E.

It can be noticed from the results that the expected AIS contacts have created flags and from those flags, risks of

TABLE I
Flags raised and risks activated for AIS contacts of the α case

Case α	1	2	3	4	5	6	7	8	9
f_nextpos					√	√			
f_trajectory									
f_ubiquity					√	√			
Collision					√	√			
Grounding									
Illegal Fishing					√	√			
Piracy / Terrorism					√	√			
Illegal Transport.					√	√			

TABLE II
Flags raised and risks activated for AIS contacts of the β case

Case β	1	2	3	4	5	6	7	8	9
f_nextpos					√				
f_trajectory			√	√	√	√			
f_ubiquity					√				
Collision			√	√	√	√			
Grounding									
Illegal Fishing			√	√	√	√			
Piracy / Terrorism			√	√	√	√			
Illegal Transport.			√	√	√	√			

TABLE III
Flags raised and risks activated for AIS contacts of the γ case

Case γ	1	2	3	4	5	6	7	8	9
f_nextpos					√	√	√	√	√
f_trajectory									
f_ubiquity									
Collision									
Grounding									
Illegal Fishing					√	√	√	√	√
Piracy / Terrorism					√	√	√	√	√
Illegal Transport.					√	√	√	√	√

TABLE IV
Flags raised and risks activated for AIS contacts of the δ case

Case δ	1	2	3	4	5	6	7	8	9
f_nextpos		√	√	√	√	√	√	√	√
f_trajectory									
f_ubiquity		√	√	√	√	√	√	√	√
Collision		√	√	√	√	√	√	√	√
Grounding									
Illegal Fishing		√	√	√	√	√	√	√	√
Piracy / Terrorism		√	√	√	√	√	√	√	√
Illegal Transport.		√	√	√	√	√	√	√	√

various nature have been individuated. Since it is a broad study, a variety of possible risks have been extracted from the expert-based rule system, but the majority of them could then be dismissed by a human operator in the case of maritime surveillance activities.

VI. Conclusion

In this paper, we presented a series of uses and misuses of the widespread Automatic Identification System, and the necessity of processing AIS data for a comprehensive knowledge of maritime traffic, both at a local and at a global scale. We proposed a methodology for message assessment based on items of data integrity. A non-genuine message is then expected to raised some computational flags linked to the integrity items, allowing a further risk analysis and the determination of specific risks linked to the presence of a vessel transmitting untrue data. An exemplification of the methodology was proposed on a series of experiments involving *ad hoc* synthetised AIS contacts simulating a variety of kinematic issues integrated into a larger AIS dataset. The determination of a variety of risks with associated risks levels is expected to support the monitoring of maritime traffic, both by ship-owners and on-shore maritime authorities, and thus enhance the level of maritime domain awareness and maritime safety and security.

ACKNOWLEDGMENT

The research presented in this paper has been supported by The French National Research Agency (ANR) and co-funded by DGA (French Directorate General of Armaments) under reference ANR-14-CE28-0028, in the frame of the DéAIS project, labelled by French clusters Pôle Mer Bretagne Atlantique and Pôle Mer Méditerranée.

REFERENCES

[1] Laurent Lecornu, Julien Montagner, and John Puentes. Reliability evaluation of incomplete AIS trajectories. In *Proceedings of the COST MOVE Workshop on Moving Objects at Sea*, June 2013.

[2] Philip A. McGillivary, Kurt D. Schwehr, and Kevin Fall. Enhancing ais to improve whale-ship collision avoidance and maritime security. In *Proceedings of the OCEANS 2009 Biloxi Conference*. IEEE, 2009.

[3] Philipp Last, Martin Hering-Bertram, and Lars Linsen. How automatic identification system (AIS) antenna setup affects AIS signal quality. *Ocean Engineering*, 100:83–89, May 2015.

[4] Abbas Harati-Mokhtari, Alan Wall, Philip Brooks, and Jin Wang. Automatic identification system (ais): Data reliability and human error implications. *Journal of Navigation*, 60(3):373–389, 2007.

[5] Xiaobo Qu, Qiang Meng, and Li Suyi. Ship collision risk assessment for the Singapore Strait. *Accident Analysis & Prevention*, 43(6):2030–2036, November 2011.

[6] Jinxian Weng, Qiang Meng, and Xiaobo Qu. Vessel Collision Frequency Estimation in the Singapore Strait. *Journal of Navigation*, 65(02):207–221, April 2012.

[7] P. Lessing, L. Bernard, B. Tetreault, and J. Chaffin. Use of the Automatic Identification System (AIS) on Autonomous Weather Buoys for Maritime Domain Awareness Applications. In *Proceedings of the OCEANS 2006 Boston Conference*, pages 1–6. IEEE, September 2006.

[8] Philippe Gaspar, Rémy Lopez, Marza Marzuki, Ronan Fablet, Philippe Gros, Jean-Michel Zigna, and Gaetan Fabritius. Analysis of vessel trajectories for maritime surveillance and fisheries management. In Michele Vespe and Fabio Mazzarella, editors, *Proceedings of the Maritime Knowledge Discovery and Anomaly Detection Workshop*, JRC Conference and Workshop Reports, pages 22–23, 2016.

[9] Michel Morel, Aldo Napoli, Anne Littaye, Marie-Pierre Gleizes, Valérie Bazin, Bernard Alhadef, Christian Scapel, Bruno Leroy, Jacques Lebrevelec, and Daniel Dejardin. *Surveillance et contrôle des activités des navires en mer*. CNRS Editions, 2009.

[10] Fabrizio M. Maggi, Arjan J. Mooij, and Wil M. P. van der Aalst. *Analyzing Vessel Behavior Using Process Mining*, pages 133–148. Springer New York, New York, USA, 2013.

[11] Fabio Mazzarella, Virginia Fernandez Arguedas, and Michele Vespe. Knowledge-based vessel position prediction using historical AIS data. In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. Institute of Electrical and Electronics Engineers (IEEE), oct 2015.

[12] B.J. Tetreault. Use of the Automatic Identification System (AIS) for Maritime Domain Awareness (MDA). In *Proceedings of OCEANS 2005 MTS/IEEE*. IEEE, 2005.

[13] Jun Min Mou, Cees van der Tak, and Han Ligteringen. Study on collision avoidance in busy waterways by using AIS data. *Ocean Engineering*, 37(5-6):483–490, April 2010.

[14] I. P. A. Stitt. AIS and Collision Avoidance a Sense of Déjà Vu. *Journal of Navigation*, 57(2):167–180, May 2004.

[15] Yang Wang, Jinfen Zhang, Xianqiao Chen, Xiumin Chu, and Xinping Yan. A spatialtemporal forensic analysis for inlandwater ship collisions using AIS data. *Safety Science*, 57:187–202, August 2013.

[16] Jakub Montewka, Tomasz Hinz, Pentti Kujala, and Jerzy Matusiak. Probability modelling of vessel collisions. *Reliability Engineering & System Safety*, 95(5):573–589, May 2010.

[17] Anne-Laure Jousselme and Giuliana Pallotta. Dissecting uncertainty-based fusion techniques for maritime anomaly detection. In *Proceedings of the 18th International Conference on Information Fusion*. IEEE, 2015.

[18] Steven Horn, Cheryl Eisler, Peter Dobias, and Joe Collins. Data requirements for anomaly detection. In Michele Vespe and Fabio Mazzarella, editors, *Proceedings of the Maritime Knowledge Discovery and Anomaly Detection Workshop*, JRC Conference and Workshop Reports, pages 52–56, 2016.

[19] Zhixian Yan, Jose Macedo, Christine Parent, and Stefano Spaccapietra. Trajectory Ontologies and Queries. *Transactions in GIS*, 12:75–91, December 2008.

[20] Gennady Andrienko, Natalia Andrienko, Christophe Claramunt, Georg Fuchs, and Cyril Ray. Visual analysis of vessel traffic safety by extracting events and orchestring interactive filters. In *In proceedings of the Workshop on Visually-supported Computational Movement Analysis (VCMA'16)*, page 7, 2016.

[21] Martin Redoutey, Eric Scotti, Christian Jensen, Cyril Ray, and Christophe Claramunt. Efficient Vessel Tracking with Accuracy Guarantees. In Michela Bertolotto, Cyril Ray, and Xiang Li, editors, *Proceedings of the W2GIS 2008 Conference: Web and Wireless Geographical Information Systems*, volume 5373, pages 140–151. Springer Berlin Heidelberg, December 2008.

[22] Fabio Mazzarella, Alfredo Alessandrini, Harm Greianus, Marlene Alvarez Alvarez, Pietro Argentieri, Domenico Nappo, and Ziemba Lukasz. Data fusion for wide-area maritime surveillance. In *Proceedings of the COST MOVE Workshop on Moving Objects at Sea*, 2013.

[23] Gregor Siegert, Paweł Banyś, and Frank Heymann. Improving the maritime trafffic situation assessment for a single target in a multisensor environment. In Michele Vespe and Fabio Mazzarella, editors, *Proceedings of the Maritime Knowledge Discovery and Anomaly Detection Workshop*, JRC Conference and Workshop Reports, pages 78–82, 2016.

[24] Fabio Mazzarella, Michele Vespe, Dimitrios Damalas, and Giacomo Osio. Discovering vessel activities at sea using ais data: Mapping of fishing footprints. In *Proceedings of the 17th International Conference on Information Fusion*, 2014.

[25] Fabrizio Natale, Maurizio Gibin, Alfredo Alessandrini, Michele Vespe, and Anton Paulrud. Mapping Fishing Effort through AIS Data. *PLOS ONE*, 10(6), June 2015.

[26] Erico N. de Souza, Kristina Boerder, Stan Matwin, and Boris Worm. Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning. *PLOS ONE*, 11(7), July 2016.

[27] Jinhai Chen, Feng Lu, and Guojun Peng. A quantitative approach for delineating principal fairways of ship passages through a strait. *Ocean Engineering*, 103:188–197, July 2015.

[28] M. Numano, H. Itoh, J. Futuko, and Y. Niwa. Simulation study on sea traffic control at an intersection utilizing information sharing with automatic identification system (ais), 2003.

[29] Bin Zheng, Jinbiao Chen, Shaosheng Xia, and Yongxing Jin. Data analysis of vessel traffic flow using clustering algorithms. In *Proceedings of the 2008 International Conference on Intelligent Computation Technology and Automation (ICICTA)*. Institute of Electrical and Electronics Engineers (IEEE), oct 2008.

[30] Kyay Mone Soe Oo, Chaojian Shi, Hu Qinyou, and Adam Weintrit. Clustering analysis and identification of marine traffic congested zones at wusongkou, shanghai. *Zeszyty Naukowe Akademii Morskiej w Gdyni*, 67:101–113, 2010.

[31] Maciej Gucma. Combination of processing methods for various simulation data sets. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 2(1):11–15, 2008.

[32] Krzysztof Naus, Artur Makar, and Jaroslaw Apanowicz. Usage ais data for analyzing ship's motion intensity. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 1(3):237–242, 2007.

[33] Nathan D. Merchant, Matthew J. Witt, Philippe Blondel, Brendan J. Godley, and George H. Smith. Assessing sound exposure from shipping in coastal waters using a single hydrophone and Automatic Identification System (AIS) data. *Marine Pollution Bulletin*, 64(7):1320–1329, July 2012.

[34] Christine Erbe, Alexander MacGillivray, and Rob Williams. Mapping cumulative noise from shipping to inform marine spatial planning. *The Journal of the Acoustical Society of America*, 132(5):423–428, 2012.

[35] Nathan D. Merchant, Enrico Pirotta, Tim R. Barton, and Paul M. Thompson. Monitoring ship noise to assess the impact of coastal developments on marine mammals. *Marine Pollution Bulletin*, 78(1-2):85–95, January 2014.

[36] Leila Hatch, Christopher Clark, Richard Merrick, Sofie Van Parijs, Dimitri Ponirakis, Kurt Schwehr, Michael Thompson, and David Wiley. Characterizing the Relative Contributions of Large Vessels to Total Ocean Noise Fields: A Case Study Using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. *Environmental Management*, 42(5):735–752, November 2008.

[37] Apollonia Miola and Biagio Ciuffo. Estimating air emissions from ships: Meta-analysis of modelling approaches and available data sources. *Atmospheric Environment*, 45(13):2242–2251, April 2011.

[38] Simon K.W. Ng, Christine Loh, Chubin Lin, Veronica Booth, Jimmy W.M. Chan, Agnes C.K. Yip, Ying Li, and Alexis K.H. Lau. Policy change driven by an AIS-assisted marine emission inventory in Hong Kong and the Pearl River Delta. *Atmospheric Environment*, 76:102–112, September 2013.

[39] P.S. Yau, S.C. Lee, James J. Corbett, Chengfeng Wang, Y. Cheng, and K.F. Ho. Estimation of exhaust emission from ocean-going vessels in Hong Kong. *Science of The Total Environment*, 431:299–306, August 2012.

[40] Su Song. Ship emissions inventory, social cost and eco-efficiency in Shanghai Yangshan port. *Atmospheric Environment*, 82:288–297, January 2014.

[41] David N. Wiley, Michael Thompson, Richard M. Pace, and Jake Levenson. Modeling speed restrictions to mitigate lethal collisions between ships and whales in the Stellwagen Bank National Marine Sanctuary, USA. *Biological Conservation*, 144(9):2377–2381, September 2011.

[42] Windward. AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea. Technical report, Windward, 2014.

[43] Fotios Katsilieris, Paolo Braca, and Stefano Coraluppi. Detection of malicious AIS position spoofing by exploiting radar information. In *Proceedings of the 16th International Conference on Information Fusion*, July 2013.

[44] Marco Balduzzi, Alessandro Pasta, and Kyle Wilhoit. A security evaluation of AIS Automated Identification System. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC'14, pages 436–445, New York, NY, USA, December 2014. ACM.

[45] Clément Iphar. *Formalisation of a data analysis environment based on anomaly detection for risk assessment Application to Maritime Domain Awareness*. PhD thesis, PSL Research University - MINES ParisTech, 2017.

[46] Cyril Ray, Richard Dreo, Elena Camossi, and Anne-Laure Jousselme. Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance (version 0.1), February 2018. Data set. Licence CC-BY-NC-SA-4.0. Zenodo. http://doi.org/10.5281/zenodo.1167595.

[47] Clément Iphar, Aldo Napoli, Cyril Ray, Erwan Alincourt, and David Brosset. Risk Analysis of falsified Automatic Identification System for the improvement of maritime traffic safety. In Tim Bedford Lesley Walls, Matthew Revie, editor, *Proceedings of the ESREL 2016 Conference*, pages 606–613. Taylor & Francis, September 2016.